GrantPanel

• SAMPLE REPORT · NSF

GrantPanel NSF Panel Review

with Recommended Revisions

Stress-test of a proposal: compliance check, solicitation alignment, three independent reviewers, three rounds of panel discussion, a consensus summary, and a recommended revision for every weakness. All generated by GrantPanel's AI review panel.

grantpanel.ai

PROPOSAL TITLE

Surface-Aware Graph Neural Networks for Data-Driven Discovery of Bimetallic Catalysts

PRINCIPAL INVESTIGATOR

Dr. Alex Morgan

INSTITUTION

Lakeshore State University

NSF PROGRAM

Designing Materials to Revolutionize and Engineer our Future (DMREF)

REQUESTED AMOUNT

$1,200,000

DURATION

36 months

PANEL CONSENSUS

Competitive

3 OF 4

The disagreement narrowed in round 3 but did not fully resolve.

Panel consensus. The proposal presents a graph-neural-network framework for catalyst discovery with an integrated synthesis-validation loop. Reviewers agreed the methodological core is sound and the PI is well qualified, but they raised consistent concerns about the rigor of the validation plan, the framing of the broader impacts, and a missing industry collaborator required by the solicitation.

FIX IN THIS ORDER

four of the panel's findings — the full list is below

01 **Missing merit-review sub-headings. →**

Add the Intellectual Merit and Broader Impacts sub-headings to the Project Summary (PAPPG II.D.2.b).

02 **Data Management Plan over the limit. →**

Trim it back to the 2-page maximum.

03 **Two citations missing from References Cited. →**

Add Lin et al. 2024 and Park & Wu 2023.

04 **Outdated biosketch format. →**

Regenerate Co-PI Dr. Singh's biosketch in the SciENcv format.

> This is the short list. The panel filed three independent reviews, argued across three rounds, and produced **five groups of recommended revisions** — including items not shown here. Each finding is tied to the section it came from, so you can check the claim against your own document.

Contents

DOCUMENT INTELLIGENCE

Document Intelligence

Proposal package inventory

READY WITH QUESTIONS

Detected documents:

- Project Summary (pages 1–2)
- Project Description (pages 3–17)
- References Cited (pages 18–21)
- Data Management Plan (2 pages) and Budget Justification (3 pages) are present in the supporting package.

Conditional requirements:

- Industry collaboration documentation: unclear. The solicitation emphasises industry participation and Section 4 names two partners, but no collaboration letter was identified in the package.
- Postdoctoral Mentoring Plan: not applicable. No postdoctoral personnel are described in the budget or personnel section.

Document quality: No document quality issues detected. Text extraction was clean across all 26 pages.

Author questions: Will the industry collaborator letters be included in the final submission package? The panel judged Broader Impacts as if they will not.

COMPLIANCE & ALIGNMENT

Compliance

NSF formatting & eligibility check

COMPLETED

Status: Compliant with concerns

Issues found:

- Project Summary is missing the explicit "Intellectual Merit" sub-heading required by PAPPG II.D.2.b — Reviewers may struggle to find the IM statement.
- Data Management Plan exceeds the 2-page limit (currently 2 pages + 4 lines onto page 3).
- Two cited references appear in the proposal text but are missing from the References Cited section ("Lin et al. 2024", "Park & Wu 2023").
- Biosketch for Co-PI Dr. Singh uses the older NSF format; PAPPG requires the NIH-style SciENcv format.

Notes: No issues with page limits on Project Description (14/15 pages) or formatting (font, margins, line spacing all compliant).

Solicitation Alignment

Does the proposal address the solicitation?

COMPLETED

Alignment: Adequate

Addressed well:

- The proposal directly addresses the solicitation's "data-driven catalysis" theme (Sec. 3.1) and references the program's stated interest in "high-throughput experimental–computational loops" (solicitation Sec. II.A).
- The required broadening-participation plan is present and concrete (HBCU partnership, Sec. 5.1).
- Clear deliverables match the solicitation's mandate for open data products (Sec. 5.2 of proposal vs. solicitation Sec. IV.C).

Gaps:

- The solicitation explicitly requires a "multi-institutional team with at least one industry collaborator" (Sec. II.B); the proposal has multi-institutional academic partners but no industry letter of collaboration.

- The mandatory cybersecurity / data-stewardship plan for the open dataset (solicitation Sec. IV.D) is not addressed.
- Workforce-development outcomes are described qualitatively; the solicitation requires "quantitative success metrics" (Sec. V.B).

INDIVIDUAL REVIEWS

Reviewer A
NSF panelist

COMPLETED

RATING — Very Good — 4 OF 5

Summary. The proposal develops a graph neural network framework for predicting catalytic activity from molecular structure, with experimental validation on a small set of bimetallic catalysts.

Intellectual Merit — strengths

- Novel message-passing scheme tailored to surface chemistry (Sec. 3.2) appears genuinely new.
- PI has strong track record on related ML-for-chemistry work (5 relevant papers in past 3 years).
- Computational and experimental components are well integrated.

Intellectual Merit — weaknesses

- The proposed validation set (n=12) is too small to support the generalization claims in Sec. 4.
- The uncertainty quantification plan (Sec. 3.4) is sketched but lacks a baseline comparison.
- No discussion of failure modes when the model is out-of-distribution.

Broader Impacts — strengths

- Plans to release an open dataset of measured activities (Sec. 5.1).
- REU site partnership with an HBCU is concrete and budgeted.

Broader Impacts — weaknesses

- Curriculum integration claims (Sec. 5.2) are aspirational with no specific courses named.

Solicitation alignment. Addresses the "data-driven discovery" priority directly; less attention to the cross-disciplinary collaboration requirement.

Overall. Strong technical core with a real weakness around validation rigor; the broader impacts are uneven.

Reviewer B
NSF panelist

COMPLETED

RATING — Good — 3 OF 5

Summary. A graph-neural-network approach to catalyst discovery, paired with limited experimental validation and a community dataset deliverable.

Intellectual Merit — strengths

- The architectural innovation (Sec. 3.2) is plausible and incremental in a good way.
- Preliminary results (Fig. 4) show the method outperforms two reasonable baselines on a held-out set.

Intellectual Merit — weaknesses

- The methodology section reads as a stack of techniques without a clear hypothesis-driven plan.
- Statistical analysis plan is absent — what counts as "improvement" is never operationalized.
- Computational cost of training the proposed model at scale is not addressed.

Broader Impacts — strengths

- Open dataset release is a tangible community contribution.

Broader Impacts — weaknesses

- Outreach plan is generic and does not differentiate this proposal from typical NSF efforts.
- No discussion of mentoring structure for the proposed REU students.

Solicitation alignment. Partial. The proposal meets the AI/ML focus but underdelivers on the solicitation's explicit call for measurable workforce-development outcomes.

Overall. Competent but not exceptional; the lack of a hypothesis-driven framing and weak BI plan hold this back.

C

Reviewer C
NSF panelist

COMPLETED

RATING — Very Good ●●●●○ 4 OF 5

Summary. The PI proposes a machine-learning method for catalyst screening combined with targeted synthesis experiments.

Intellectual Merit — strengths

- Excellent fit between the PI's prior work and the proposed direction.
- The experimental loop (Sec. 4.3) is realistic and the timeline is credible.
- The ablation studies in the preliminary data are unusually thorough for a proposal at this stage.

Intellectual Merit — weaknesses

- The "transferability across reaction classes" claim (Sec. 2) is not supported by any preliminary evidence.
- Risk-mitigation discussion is brief and does not address the most plausible failure mode (data scarcity).

Broader Impacts — strengths

- Concrete HBCU partnership with named collaborator and committed funding.
- Open-source code and dataset releases are well specified.
- Plans for a public-facing tutorial at a major conference.

Broader Impacts — weaknesses

- Workforce-development metrics (Sec. 5.3) are vague.

Solicitation alignment. Strong alignment with both the AI/ML priority and the broadening-participation requirement.

Overall. A well-constructed proposal with a tight experimental plan and stronger-than-average broader impacts; the transferability claim is the main concern.

PANEL DISCUSSION

ROUND 1 — POSITIONS

A — **REVIEWER A**
I'm landing at Very Good. The technical contribution is real but the n=12 validation set is the central concern for me.

B — **REVIEWER B**
I'm at Good. The methodology is a stack of techniques without a hypothesis-driven plan, and the BI section is generic outside the HBCU piece.

C — **REVIEWER C**
Very Good. The preliminary ablations and the HBCU partnership are unusually strong; the transferability claim is my main concern.

ROUND 2 — COUNTER-POINTS

A — **REVIEWER A**
Reviewer B, I take the point on framing, but the proposal does articulate a measurable scientific question in Sec. 2.4. I'd push back on calling it purely methods-first.

B — **REVIEWER B**
Fair on Sec. 2.4, but it's still a methods proposal in disguise. And the compliance flag on the missing industry collaborator is a solicitation-level issue we can't hand-wave away.

C

REVIEWER C

The industry-collaborator gap is real, but the rest of the alignment is strong. I don't think it should sink an otherwise solid proposal — though it should be communicated to the PI.

ROUND 3 — FINAL VOTES

A

REVIEWER A

Final rating: Very Good. The PI's record and the experimental rigor outweigh the validation-size concern for me.

B

REVIEWER B

Final rating: Good. The framing weakness, the BI generic-ness, and the solicitation gap keep me below Very Good.

C

REVIEWER C

Final rating: Very Good. Strong preliminary data and a concrete BI partnership; transferability is a flag for the PI but not disqualifying.

PANEL SUMMARY

Panel Summary

Competitive 3 OF 4

Panel consensus. The proposal presents a graph-neural-network framework for catalyst discovery with an integrated synthesis-validation loop. Reviewers agreed the methodological core is sound and the PI is well qualified, but they raised consistent concerns about the rigor of the validation plan, the framing of the broader impacts, and a missing industry collaborator required by the solicitation.

Intellectual Merit — Consensus strengths

- The PI has clear domain expertise and a strong publication record in ML-for-chemistry.
- The proposed message-passing architecture is a plausible technical contribution.
- Preliminary results, particularly the ablations and baselines, exceed typical proposal-stage rigor.
- The synthesis-validation loop is well integrated with the computational plan.

Intellectual Merit — Consensus weaknesses

- The validation set is too small (n=12) to support the generalization claims; reviewers agreed this is the central technical risk.
- The transferability-across-reaction-classes claim lacks preliminary evidence.
- The statistical analysis plan is not operationalized — there is no a-priori definition of success.
- Risk mitigation does not address the most plausible failure mode (data scarcity).

Broader Impacts — Consensus strengths

- The HBCU partnership for the REU component is concrete, named, and budgeted.
- The open dataset and code release are tangible community contributions.
- A public-facing tutorial at a major conference is planned.

Broader Impacts — Consensus weaknesses

- Outreach beyond the HBCU partnership is generic and does not differentiate this proposal from typical NSF efforts.
- No quantitative workforce-development metrics are specified.
- Mentoring structure for proposed REU students is not described.

Points of disagreement

- Rating split (Good vs. Very Good): Reviewer B held the more critical view based on weak hypothesis framing and the generic BI plan; A and C placed more weight on the experimental rigor and the REU partnership. The disagreement narrowed in round 3 but did not fully resolve.
- Weight of the missing industry collaborator: B treated this as a solicitation-level disqualifier; C considered it a communicable concern rather than a sinker.

Justification. The proposal has a credible technical core, a strong PI, and at least one stand-out broader-impacts element, but the validation rigor, the missing industry collaborator (a solicitation requirement), and the absence of quantitative BI metrics prevent it from reaching the Highly Competitive tier. The panel encourages the program officer to consider funding pending available resources and to communicate the validation and collaborator concerns to the PI.

RECOMMENDED REVISIONS

Every weakness and gap the panel raised, paired with a specific, actionable revision, and grouped by review stage.
GrantPanel recommends changes and shows you where; it never edits your document.

Revisions by section

Ordered by what to fix first

5 REVISIONS

01 Compliance & formatting

FIX BEFORE SUBMISSION

→ **Missing merit-review sub-headings.** Add the *Intellectual Merit* and *Broader Impacts* sub-headings to the Project Summary (PAPPG II.D.2.b).

→ **Data Management Plan over the limit.** Trim it back to the 2-page maximum.

→ **Two citations missing from References Cited.** Add Lin et al. 2024 and Park & Wu 2023.

→ **Outdated biosketch format.** Regenerate Co-PI Dr. Singh's biosketch in the SciENcv format.

02 Solicitation alignment

MEET REQUIREMENTS

→ **No industry collaborator.** Secure an industry letter of collaboration to satisfy the required multi-institutional team.

→ **Missing data-stewardship plan.** Add the cybersecurity / data-stewardship plan required for the open dataset (Sec. IV.D).

→ **Qualitative workforce outcomes.** Convert workforce-development outcomes into the quantitative success metrics the solicitation requires.

03 Individual reviews

STRENGTHEN THE SCIENCE

INTELLECTUAL MERIT

→ **Validation set too small (n=12).** Expand it, or add a held-out benchmark with cross-validation, plus an out-of-distribution failure-mode analysis.

→ **Uncertainty-quantification plan lacks a baseline.** Add a baseline comparison for the UQ method.

→ **Methodology reads as a stack of techniques.** Reframe around a testable hypothesis and add a statistical analysis plan with an a-priori success criterion.

→ **Training compute not addressed.** Provide a training-cost estimate at scale with a mitigation strategy.

→ **Transferability claim unsupported.** Back the transferability-across-reaction-classes claim with preliminary data, or scope it to the tested classes; add data-scarcity risk mitigation.

BROADER IMPACTS

→ **Curriculum integration is aspirational.** Name specific courses and terms, with a supporting letter or draft syllabus.

→ **Generic outreach, no mentoring plan.** Replace with project-specific activities and define an REU mentoring structure (ratios, training, milestones).

→ **Vague workforce metrics.** Specify number of trainees, recruitment targets, and measurable outcomes.

04 Panel consensus

ADDRESS FIRST

→ **Highest priority.** Validation rigor (the n=12 set) and the missing industry collaborator — the two issues most likely to gate a fundable outcome.

→ **Then.** Operationalize the statistical analysis plan and add data-scarcity risk mitigation.

→ **Broader impacts.** Quantify the workforce-development metrics and describe the REU mentoring structure.

05 Prior work & citations

POSITION THE CONTRIBUTION

→ **Four directly relevant works uncited.** Cite and engage Riebesell (Matbench Discovery 2024), Tran & Ulissi (Nat. Catal. 2024), Liao & Smidt (EquiformerV2, ICLR 2024), and the Sun group closed-loop study (Sci. Adv. 2023).

SUPPORTING RESEARCH

REFERENCES LOOKUP

References
Extracts and looks up cited prior work

COMPLETED

Extracting cited references from the proposal...

Found 6 references. Fetching abstracts from Semantic Scholar...

[1/6] Graph neural networks for materials discovery — found [2/6] Bayesian uncertainty quantification for deep ensembles — found [3/6] High-throughput experimental synthesis of bimetallic catalysts — found [4/6] Closed-loop autonomous chemistry — found [5/6] Open data infrastructure for catalysis — no match [6/6] REU programs and undergraduate research outcomes — found

---

The proposal cites 6 references. The abstract or TLDR of each (when available) is provided below to inform your assessment of novelty and prior-work coverage. Cited works without a match in Semantic Scholar are listed by their proposal entry only.

[1] Graph neural networks for materials discovery *(2023)* — Lin et al. *Nature Machine Intelligence · 2023* We present a message-passing graph neural network for predicting the energetic and electronic properties of inorganic materials. Our architecture incorporates explicit surface-chemistry inductive biases and achieves state-of-the-art accuracy on the OQMD benchmark while requiring 40% fewer parameters than competing models.

[2] Bayesian uncertainty quantification for deep ensembles *(2022)* — Park & Wu *NeurIPS · 2022* Deep ensembles provide a simple and effective approach to uncertainty quantification but lack a principled Bayesian interpretation. We derive a connection between ensembles and approximate Bayesian inference and show that calibrated uncertainty estimates emerge when ensemble members are trained with mild functional diversity.

[3] High-throughput experimental synthesis of bimetallic catalysts *(2024)* — Chen et al. *Journal of Catalysis · 2024* We describe an automated parallel synthesis platform capable of producing and characterizing 192 bimetallic catalyst compositions per week. The platform integrates colloidal synthesis, XRD, and reactor screening, providing the throughput necessary for data-driven catalyst discovery.

[4] Closed-loop autonomous chemistry *(2022)* — Burger et al. *Nature · 2022* We demonstrate a mobile robotic chemist that conducts autonomous photocatalytic hydrogen evolution experiments using Bayesian optimization. Over 8 days the system identified a catalyst formulation six times more active than the human-chosen baseline.

[5] Open data infrastructure for catalysis *(Not found on Semantic Scholar.)*

[6] REU programs and undergraduate research outcomes *(2021)* — Russell et al. *Science · 2021* A longitudinal study of 3,400 participants in NSF REU programs found significant increases in students' research self-efficacy, identification as a scientist, and probability of pursuing a PhD, with the largest effects observed for students from groups historically underrepresented in STEM.

LITERATURE REVIEW (DEEP RESEARCH)

Literature Review
Deep-research SOTA scan

COMPLETED

Research area & central claim. The proposal develops a graph neural network framework for catalyst discovery, integrating a novel surface-chemistry-aware message-passing scheme with a closed-loop experimental validation platform for bimetallic catalysts. The central claim is that this combination will produce calibrated, transferable predictions of catalytic activity that outperform existing benchmarks while operating with smaller training sets.

State of the art (last 2-3 years)

- Burger et al. (Nature, 2022) demonstrated autonomous robotic chemistry with Bayesian optimization, identifying a photocatalyst 6× more active than human-chosen baselines over 8 days.
- Choudhary & DeCost (npj Comput. Mater., 2023) released the JARVIS-DFT benchmark and ALIGNN architecture, setting a recent SOTA on materials property prediction across multiple inorganic property targets.
- Lin et al. (Nature Machine Intelligence, 2023) introduced surface-chemistry-aware GNNs for adsorption energy prediction, with explicit symmetry-equivariant message passing.
- Chen et al. (Journal of Catalysis, 2024) reported a 192-composition-per-week parallel synthesis platform for bimetallic catalysts integrated with XRD and reactor screening.
- Schwaller et al. (Chem. Sci., 2023) demonstrated transferable reaction yield prediction with attention-based encoders trained on cross-domain heterogeneous catalysis data.
- Tran & Ulissi (Nat. Catal., 2024) showed active-learning loops that reduced DFT calls by ~80% in catalyst screening.
- Riebesell et al. (Matbench Discovery 2024) reported that state-of-the-art models still show 15-25% out-of-distribution error on bimetallic systems, highlighting an open challenge.

Competing approaches

- **MatGNN family** (Choudhary, Karpur, Anubhav Jain) — strong on bulk crystal properties; weaker on surface chemistry, which the proposal targets.
- **ULM-Mol / EquiformerV2** (Liao, Smidt) — equivariant transformers achieving SOTA on OC20; more parameters than the proposal's approach but no integrated experimental loop.
- **Chemspeed / Emerald Cloud Lab platforms** — commercial autonomous-chemistry stacks; faster throughput but proprietary, not open to community benchmarking.
- **Sun group (Toronto)** — closed-loop electrocatalyst discovery with Bayesian optimization; the proposal's loop is similar in spirit but uses GNN priors rather than Gaussian processes.

Open problems the proposal addresses

- Out-of-distribution generalization across reaction classes — a recognized failure mode of current bimetallic catalyst GNNs (Riebesell 2024).
- Uncertainty calibration in low-data regimes — most published work does not provide reliable predictive intervals at the small-N validation scales the proposal proposes.
- Integration of GNN priors with experimental loops — work to date uses GPs or random forests; the proposal's GNN-driven loop is comparatively new.

Citation gaps in the proposal

- Riebesell et al., Matbench Discovery 2024 — directly relevant for out-of-distribution evaluation methodology; not cited.
- Tran & Ulissi, Nat. Catal. 2024 — the most-cited recent active-learning result for catalyst screening; not cited.
- Liao & Smidt, EquiformerV2 (ICLR 2024) — an equivariant baseline the proposal should compare against; not cited.
- Sun group's closed-loop work (Sci. Adv. 2023) — directly comparable on the experimental-loop side; not cited.

Search log

- "graph neural network bimetallic catalyst activity 2024"
- "autonomous catalyst discovery closed loop"
- "Matbench Discovery out of distribution bimetallic"
- "equivariant GNN materials property benchmark"
- "active learning DFT catalyst screening 2023 2024"
- "uncertainty quantification deep ensembles materials"
- "high throughput bimetallic catalyst synthesis 2024"
- "Sun group autonomous electrocatalyst"
- "Lin surface chemistry message passing GNN"
- "Tran Ulissi active learning catalysis 2024"
- "JARVIS DFT benchmark ALIGNN"
- "open dataset catalysis community 2024"
- "REU NSF computational catalysis outcomes"
- "Bayesian uncertainty calibration graph neural networks"
- "EquiformerV2 OC20 benchmark"

INDIVIDUAL WEB SEARCHES

Reviewer A
3 searches

1. `graph neural network bimetallic catalyst activity 2024`
2. `n=12 validation set generalization machine learning chemistry`
3. `uncertainty quantification deep ensembles materials science 2024`

Reviewer B
4 searches

1. `hypothesis-driven NSF proposal framing best practices`
2. `statistical analysis plan machine learning preregistration`
3. `Matbench Discovery out of distribution 2024`
4. `REU mentoring structure NSF best practices`

Reviewer C
3 searches

1. `transferability across reaction classes catalyst GNN`
2. `HBCU REU partnership outcomes computational chemistry`
3. `EquiformerV2 OC20 baseline comparison`